SETSCI conference proceedings
SETSCI Conference Proceedings
4 (6), 497-501, 2019

4th International Symposium on Innovative Approaches in Engineering and Natural Sciences
November 22-24, 2019, Samsun, Turkey

https://doi.org/10.36287/setsci.4.6.139
2618-6446/ © 2019 The Authors. Published by SETSCI

ISAS WINTER - 2019

Akut Lenfoblastik Löseminin Makine Öğrenimi ile Sınıflandırılması

Ayşe Berika Varol Malkoçoğlu [2+], İsmail İşeri[*]

[1]*Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü, Ondokuz Mayıs Üniversitesi, Samsun, Türkiye*
[2] *Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü, Maltepe Üniversitesi, İstanbul, Türkiye*
[*]*Danışman yazar: ismail.iseri@bil.omu.edu.tr*
[+]*Konuşmacı: ayseberikavarol@maltepe.edu.tr*
Sunum/Bildiri Tipi: Oral / Tam Metin

Özet – Akut Lenfoblastik Lösemi (ALL) genellikle 2-5 yaş arasındaki çocuklarda görülen, kemik iliğindeki lenfoblastların hızlı artışından ve şekilsel bozukluklarından kaynaklanan bir kanser türüdür. Tespit edilip tedavi sürecinin başlatılmadığı takdirde ölümlere yol açabilmekte olan bu hücrelerin normal hücrelere morfolojik açıdan benzemesi, hastalığın teşhisini zorlaştırdığı bilinmektedir. Bu çalışmada Naive Bayes (NB), Random Forest (RF) ve Destek Vektör Makinesi (Support Vector Machine – SVM) algoritmalarından yararlanılarak hücrelerin (kanserli-kansersiz) sınıflandırması yapılmıştır. Sınıflandırma için açık kaynaklı Cancer Imaging Archive tarafından sunulan ALL hastalığına ilişkin 101 hastadan alınmış 12528 adet patoloji görüntüsü kullanılmıştır. Bu görüntülerin 8491 tanesi kanserli, 4037 tanesi kansersiz olarak etiketlenmiştir. ALL veri seti kullanılarak RGB (Red Green Blue) ve Gray Scale renk uzaylarında sınıflandırma performansları test edilmiştir. Test sırasında her iki renk formatı içinde orijinal ve arttırılmış veri kullanılarak 2000, 6500, 10320, 12130 büyüklüğünde hastalıklı & sağlıklı hücrelerin oluşturduğu 4 farklı veri kümesi kullanılarak, her bir görüntünün 32x32 ve 128x128 ölçülerine yeniden boyutlandırılıp sınıflandırma işlemi gerçekleştirilmiştir. Yapılan deneysel çalışmalar sonucunda RGB renk uzayında 2000 boyutlu veri seti ve 128x128 boyutlarına ölçeklendirilmiş görüntüler üzerinde SVM algoritmasının %89'luk doğruluk oranı ile en iyi sonucu verdiği gözlenmiştir.

Anahtar Kelimeler– *Naive Bayes, Random Forest, Destek Vektör Makinesi, Sınıflandırma, Makine Öğrenimi,Dijital Patoloji*

Classification of Acute Lymphoblastic Leukemia Using Machine Learning

Abstract – Acute Lymphoblastic Leukemia (ALL) is a type of cancer that usually occurs in children aged 2-5 years, resulting from the rapid increase of lymphoblasts in the bone marrow and the deformities. Morphological resemblance of these cells to normal cells is known to complicate the diagnosis of the disease, if not detected and not the treatment process started, this can lead to death. In this study, Naive Bayes (NB), Random Forest (RF) and Support Vector Machine (SVM) algorithms are used to classify the cells (malign-benign). For classification, 12528 pathology images from 101 patients with ALL, presented by the open source Cancer Imaging Archive, were used. 8491 of these images were labeled as malign and 4037 of them were labeled as benign. Using ALL data set, classification performances in RGB (Red Green Blue) and Gray Scale color spaces were tested. During the test, the original and increased data in both color formats using 2000, 6500, 10320, 12130 size of malign & benign cells using 4 different data sets, each image was resized to 32x32 and 128x128 size classification was performed. As a result of the experimental studies, it was observed that the SVM algorithm, which was run on the 2000 dimensional dataset and the images scaled to 128x128 dimensions, gave the best result with 89% accuracy ratio.

Keywords – Naive Bayes, Random Forest, Support Vector Machine, Classification, Machine Learning, Digital Pathology

I. GİRİŞ

Lösemi dünya genelinde çocuklarda en çok görülen kanser türlerinden biri olarak bilinmektedir. Fig.1'de bulunan 2018 yılına ait kanser istatistiklerine bakıldığında, dünyadaki 0-14 yaş arasındaki kanserli çocukların lösemi türüne sahip olma riskinin yüksek olduğu görülmektedir. Fig.2'ye bakıldığında ise bu oranın ülkemizde daha yüksek olduğu görülmektedir [1].

Akut Lenfoblastik Lösemi (ALL) çocuklarda görülen löseminin %75-80'ini oluşturmaktadır. Genellikle 2-5 yaş arasındaki çocuklarda görüldüğü ve hastalığın hızlı ilerlediği bilinmektedir. ALL, lenfoblastların (olgunlaşmamış lenfler) kontrolsüz ve hızlı artışı sonucunda normal kan hücrelerinin görevlerini yerine getirememisinden kaynaklanan bir kan kanseri türüdür [2], [3]. Kanserin morfolojik açıdan incelenip diğer hücrelerden ayrıştırılması ve tespit edilmesinin zor olduğu bazı çalışmalarda bahsedilmektedir. Tespitin zor olmasının sebebi lenfoblastların lenfositlere şekilsel benzerlikleri ve küçük farklılıklarından kaynaklanmaktadır. Lenfositler düzenli kenarı olan yoğun çekirdeğe sahipken, lenfoblatlar düzensiz, parçacıklı çekirdeğe ve sitoplazma ile arasında boşluğa sahip bir yapıdadır. Bu yöntemle hastalığın insan tarafından fark edilip teşhis edilebilmesi oldukça zordur. Teşhis için görüntülerin bilgisayar tabanlı sistemlere verilmesi daha nesnel ve doğru sonuçlar elde etmemizi sağlayabilmektedir [4].

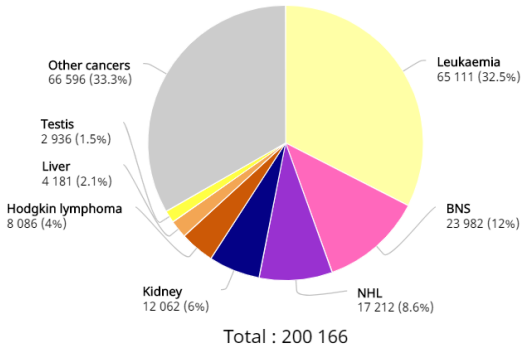


Fig.1 2018 yılı 0-14 yaş arası dünya kanser istatistikleri WHO tarafından paylaşılmıştır.

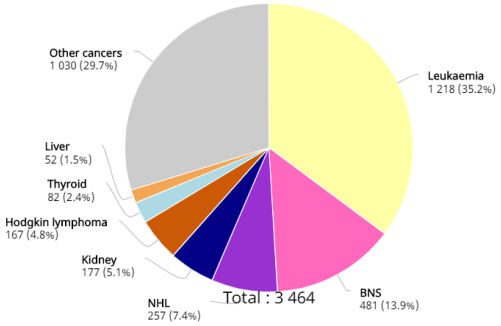


Fig.2 2018 yılı 0-14 yaş arası Türkiye kanser istatistikleri WHO tarafından paylaşılmıştır.

Abdeldaim ve diğerleri [4] 130 sağlıklı 130 kanserli kan hücresi görüntülerinden oluşan ALL-IDB2 veri setini kullanarak hücreleri kanserli/kansersiz olarak sınıflandırabilmek için sistem geliştirmişlerdir. İlk olarak görüntü görüntüleri RGB den CMYK (Cyan, Magenta, Yellow, Key) renk formatına dönüştürüp histogram eşitleme ve eşikleme işlemlerini uygulayarak görüntünün bölütlemişlerdir. Ardından şekil, renk özelliklerine bakıp normalize ederek görüntünün özelliklerini çıkarmışlardır. Son olarak da KNN, SVM, NB ve DT algoritmalarıyla sınıflandırma işlemini gerçekleştirerek hücreleri kanserli ve normal olarak sınıflandırmayı başarmışlardır. Farklı algoritmaları kullanarak performans karşılaştırması yapmışlardır. Sonuç olarak KNN algoritması bu sistemde farklı özellikleri kullanarak yapılan deneylerin sonucunda %86-96 aralığında her özellik kombinasyonunda en iyi sonucu veren algoritma olduğunu göstermişlerdir.

Bhuiyan ve diğerleri [5], kan hücreleri görüntüsüne sahip ALL-IDB1 veri tabanından alınan 108 adet görüntü verisini kullanarak RF, SVM, Lojistik Regresyon (Logistic Regression -LR) ve Karar Ağacı (Desicion Tree - DT) algoritmaları ile sınıflandırıp çeşitli metriklere göre performans karşılaştırması yapmışlardır. Sınıflandırma işleminden önce verileri kırparak beyaz kan hücrelerini elde etmişlerdir. hücrenin patlayan/patlamayan ayrımı için renk filtrelemesi yapmışlardır. Kenar bulma işleminden sonra HOG (Histogram of Oriented Gradients) özellik tanımlayıcısı, normalleştirme ve son olarak boyut küçültme işlemini yapmışlardır. Böylece özellik çıkarımını gerçekleştirip seçtikleri sınıflandırıcılara vererek sınıflandırmışlardır. Sınıflandırmanın sonucunda SVM algoritması %99.05 ACC, %99,2 Precision, %99 Recall ve %99 F-score oranıyla en başarılı algoritma olduğunu belirtmişlerdir.

Selçuk ve diğerleri [6], 130 adet ALL hastalığına sahip, 130 adet sağlıklı hücreden oluşan veri kümesi özellik çıkarımı için dalgacık dönüşümünden geçirmişlerdir. Ardından görüntünün renk skalalarını belirleyerek uygun olmayan hücreleri gri seviyeye çevirmişlerdir. Verileri 140 âdetini eğitim 120 âdetini test için kullanarak SVM algoritmasında sınıflandırma işlemini gerçekleştirmişlerdir. 10 iterasyon sonucunda %93,91 ACC başarı oranı elde etmiştir.

Bu çalışmada kan hücrelerinden alınan örneklerin kanserli/kansersiz sınıflandırması yapılmıştır. 2. Bölümde makine öğreniminden, 3. Bölümde yapılan çalışmadaki veri seti, çalışılan platform ve yöntemler anlatılırken, 3. Bölümde elde edilen bulgular 4. Bölümde ise elde edilen sonuçlar ve gelecekteki çalışmalar anlatılmaktadır.

II. MATERYAL E YÖNTEM

A. *Veri Seti*

Cancer Imaging Archive tarafından sunulan ALL veri seti toplamda 6500 adet veri içerecek şekilde yeniden düzenlenmiştir. Bu verilerin hepsi etiketli olup 3250 tanesi sağlıklı, 3250 tanesi kanserli hücrelerden oluşan görüntüleri içermektedir. Veriler 450x450 görüntü ölçeğinde sunulmuştur. Alınan veri seti kanserli ve kansersiz görüntüler eşit olacak şekilde 2000, 6500, 10320 ve 12130 adet veri içeren setler halinde 4 grup oluşturulmuştur.

10320 ve 12130 adet veri içerek kümelere veri arttırma işlemi uygulanmıştır. Arttırma için 4800 adet veri içeren yeni bir küme oluşturularak buradaki görüntüler sırasıyla 2 ve 3 kez 30 derece döndürülmüş parlaklığı 0.1 oranında arttırılmış ve 0.1 oranında kaydırılarak yeni görüntüler elde edilmiştir. Bu veriler cross-validation yöntemi ile eğitim ve validasyon olarak bölünmüştür. Algoritmaların başarı oranlarını değerlendirmek için ise ağın daha önceden hiç görmediği 200 veriden oluşan yeni bir test veri seti kullanılmıştır.

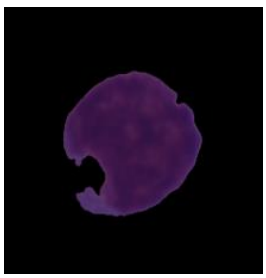


Fig.3 Kanserli kan hücresi



Fig.4 Kansersiz kan hücresi

B. *Kullanılan Donanım Platformu*

Geliştirilen uygulama için Keras, OpenCV, Sklearn, TensorFlow, Numpy, Matplotlib gibi birçok kütüphaneyi destekleyen ve ücretsiz GPU desteği sağlayan Google Colab platformu kullanılmıştır. Fig.4'de görüldüğü gibi kullanılan platformun Tesla K 80 GPU, Intel Xeon CPU 2.3 GHz, 2 çekirdek, 12 gb RAM ve 311 GB disk kapasitesi gibi fiziksel özelliklere sahip olduğu görülmektedir.

Uygulama Pyhon dilinde yazılmıştır. Colaba verileri çekebilmek için veriler öncelikle Google Drive'a yüklenmiştir. Google Drive ile Colab arasında bağlantı

kurularak verilere erişim sağlanmıştır. Verileri işleyebilmek ve üzerinde işlem yapıp sınıflandırabilmek için Numpy, Matplotlib ve Sklearn kütüphanelerinden yararlanılmıştır.

C. Kütüphaneler

- *NumPy:* N boyutlu diziler, karmaşık fonksiyonlar, matematiksel dönüşümler, mantıksal ve istatiksel işlemler, sıralama, seçme gibi sayısal işlemlerde kullanılan temel Python kütüphanesidir [7].
- *Matplotlib:* 2 boyutlu görsel işlemlerinde kullanılan Python kütüphanesidir. Görüntü verilerini kullanabilmek, ekrana yansıtabilmek için ya da sonuçları grafik, histogram, çizelge şeklinde görselleştirme için kullanılan çizim kütüphanedir [8].
- *Sklearn:* Açılımı Skit-Learn şeklin olan Python kütüphanesidir. Veri madenciliğinde etkin bir araç olan bu kütüphane sınıflandırma, kümeleme, boyut küçültme, ön işleme, regresyon ve model seçimi işlemlerinde kullanılmaktadır [9].

D. Yöntem

Çalışma kapsamında makine öğrenimi algoritmalarından RF, NB ve DVM yöntemleri kullanılmıştır. Bu yöntemlerden;

- *Random Forest:* Birden fazla karar ağacının birleşmesiyle oluşan algoritmadır. Oy çokluğuna göre tahmin yapmaktadır. Gelen verinin sınıfını belirleyebilmek için her ağacın tahmin sonucu sayılmak suretiyle, oyu en çok alan sınıf verinin ait olduğu sınıf olarak belirlenmektedir. Bu sınıflandırıcının genelleme hatası topluluktaki her bir ağacın gücüne ve aralarındaki korelasyona bağlı olarak değişmektedir [10].

$Eğitim\ seti \rightarrow S \coloneqq (x_1, x_1), \dots, (x_n, x_n),$
$Özellikler \rightarrow F,$
$Ormandaki\ ağaçların\ sayısı \rightarrow B$

function RandomForest (S, F)
  $H \leftarrow \emptyset$
  for $i \in 1, \dots, B$ *do*
    $S^{(i)} \leftarrow$ S
    $h_i \leftarrow$ RandomizedTreeLearn $(S^{(i)}, F)$
    $H \leftarrow H \cup \{h_i\}$
  end for
  return H
end function

function RandomizedTreeLearn (S, F)
  $\forall$ *düğüm için*:
    $f \leftarrow F'nin\ en\ küçük\ alt\ kümesi$
    $En\ iyi$ özelliği f *olarak ayır*
  return *Eğitilen ağaç*
end function

Fig.5 Random Forest algoritmasının sözde kodu [11]

- *Naive Bayes:* Birbirinden bağımsız veriler üzerinde hızlı çalışıp başarılı sonuçlar üreten istatiksel sınıflandırıcıdır. Tembel öğrenme algoritması olarak da bilinmektedir. Sınıflandırma işlemi Bayes teoremine dayanmaktadır. Sınıflandırma yapabilmek için bir olayın tüm olasılığını hesaplamaktadır. En yüksek değere sahip duruma göre verinin sınıfını belirlemektedir [12].

$Eğitim\ seti \rightarrow T\ (t_1, t_2, \dots, t_n),$
$Tahmin\ sonuçları \rightarrow R = (r_1, r_2, \dots, r_n),$
$Sınıf \rightarrow C\ (c_1, c_2, \dots, c_n),$

function NaiveBayes (C, R):
  $P\ (C_i | R)$:
  $$P\ (C_i | R) = \frac{P(R|C_i)P(C_i)}{P(R)}$$
  $P(R|C_i) = P(r_1, r_2, \dots, r_n | C_i)$
return $C_{max} = argmax_{C_i}(R|C_i)$

Fig.6 Naive Bayes algoritmasının sözde kodu

- *Destek Vektör Makinesi:* Sonsuz boyutlu alanda doğrusal veya doğrusal olmayan verileri belirgin bir şekilde sınıflandıran algoritmadır. Temelde doğrusal sınıflandırma yapabilen bu algoritma doğrusal olmayan verileri sınıflandırabilmek için çekirdek hilesine başvurarak yeni bir düzlem oluşturmaktadır. Bu işlem sayesinde verileri birbirinden ayırarak sınıflandırılabilmektedir [13], [14].

$Eğitim\ seti \rightarrow X\ (x_1, x_2, \dots, x_n),$
$Eşik\ değeri \rightarrow t,$
$Vektör \rightarrow v$

function DestekVektörMakinesi :
  $h(x) = v^t x + t = 0$
  if $h(x) \geq +1$
    $h(x) \rightarrow Sınıf\ 1$
  else if $h(x) \leq -1$
    $h(x) \rightarrow Sınıf\ 2$
return

Fig.7 Linear SVM algoritmasının sözde kodu

Çalışma için hazırlanan her bir veri seti Fig.8 ve Fig.9 da göründüğü gibi RGB ve GrayScale renk formatlarına dönüştürülmüştür.

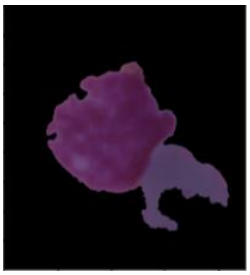


Fig.8 RGB renk uzayında kanserli hücre görüntüsü

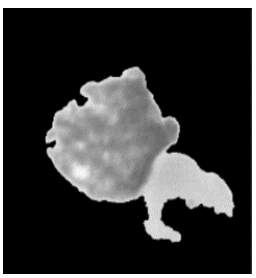


Fig.9 GrayScale renk uzayında kanserli hücre

Ardından her bir renk formatı için görüntüler 32x32 ve 128x128 boyutlarına yeniden ölçeklendirilmiştir. Fig.10 da 450x450 ölçeğindeki orijinal görüntünün diğer boyutlara ölçeklenmiş hali gösterilmektedir.

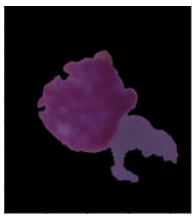


Fig.10 450x450

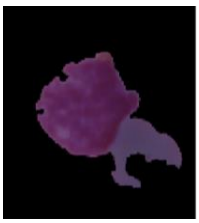


Fig.11 128x128



Fig.12 32x32

Renk formatı ve ölçekleme işlemlerinden sonra veriler karıştırılarak %70'i eğitim %30'u validasyon olacak şekilde bölünmüşlerdir. Bölünen veriler eğitim için gönderilirken test için 100 adet kanserli 100 adet kansersiz veriden oluşan farklı bir veri seti kullanılmıştır. Veriler tamamen hazırlandıktan sonra Random Forest (RF), Naive Bayes (NB) ve SVM algoritmaları kullanılarak sınıflandırma işlemi yapılmıştır. Sınıflandırma işlemi sırasında algoritmalarda çeşitli parametreler kullanılmıştır. Her bir algoritma için kullanılan parametreler şu şekildedir;

- RF algoritması için ormanda bulunan ağaçların sayısı 10 olarak belirlenmiştir. Her düğüm için seçilen rastgele özelliklerin boyutu yani maksimum özellik sayısı da 8 olarak belirlenmiştir.
- NB algoritmasında MultinominalNB modeli kullanılmıştır. Alpha özelliği 1.0, fit_prior özelliği *True* olarak belirlenmiştir böylece sınıfların önceki olasılıkları hesaba katılmıştır.
- SVM algoritmasında veriyi daha iyi haritalamak için çekirdek özelliği *linear* olarak belirlenmiştir. Her bir eğitim örneğinin hatalı sınıflandırılmasını önlemek için ise C değeri 0.01 olarak belirlenmiştir.

Algoritmaların başarı oranları seçilen parametrelere göre değişmektedir. Farklı parametreler deneyerek hangi parametre ile maksimum başarı oranı elde edildiği tespit edilmiştir.

III. BULGULAR VE TARTIŞMA

Çalışmanın test aşamasında farklı veri seti boyutları, görüntü ölçekleri ve renk formatları kullanılarak RF, NB, SVM algoritmaları üzerinde çalıştırılmıştır. Yapılan testlerin Accuracy (Doğruluk), Precision (Hassasiyet), Recall (Hatırlatma) ve F1-Score performans metrikleri sonuçları tablolarda gösterilmektedir. Metrikler tabloda gösterilirken Accuracy için Acc, Precision için Prec, Recall için Rec ve F1-Score için F1-Sc kısaltılmaları kullanılmıştır.

Tablo 1. RF algoritmasının Gray renk formatında performans sonuçları

Veri Seti Boyutu	Görüntü Ölçeği	RF			
		Acc(%)	Prec(%)	Rec(%)	F1-Sc(%)
2000	32x32	82	87	82	81
	128x128	83	86	83	83
6500	32x32	83	85	83	83
	128x128	84	88	84	84
10320	32x32	61,5	62	61	61
	128x128	69	72	69	68
12130	32x32	67	78	68	64
	128x128	46,5	46	47	46

Tablo 2. NB algoritmasının Gray renk formatında performans sonuçları

Veri Seti Boyutu	Görüntü Ölçeği	SVM			
		Acc(%)	Prec(%)	Rec(%)	F1-Sc(%)
2000	32x32	81	83	81	81
	128x128	81,5	84	81	81
6500	32x32	83	85	83	83
	128x128	83,5	85	83	83
10320	32x32	80,5	83	81	80
	128x128	81	84	81	81
12130	32x32	80,5	85	81	80
	128x128	80	84	80	79

Tablo 3.SVM algoritmasının Gray renk formatında performans sonuçları

Veri Seti Boyutu	Görüntü Ölçeği	SVM			
		Acc(%)	Prec(%)	Rec(%)	F1-Sc(%)
2000	32x32	74,5	75	74	74
	128x128	79,5	80	80	79
6500	32x32	80,5	84	81	80
	128x128	76	76	76	6
10320	32x32	75,5	81	76	74
	128x128	68	71	68	67
12130	32x32	72,5	78	72	71
	128x128	60	62	60	58

Tablo 4. RF algoritmalarının RGB renk formatında performans sonuçları

Veri Seti Boyutu	Görüntü Ölçeği	RF			
		Acc(%)	Prec(%)	Rec(%)	F1-Sc(%)
2000	32x32	84	87	84	84
	128x128	86,5	89	86	86
6500	32x32	87,5	90	88	87
	128x128	84,5	87	84	84
10320	32x32	60	63	60	58
	128x128	65	67	65	64
12130	32x32	36	21	36	26
	128x128	60	66	60	56

Tablo 5. SVM algoritmalarının RGB renk formatında performans sonuçları

Veri Seti Boyutu	Görüntü Ölçeği	NB			
		Acc(%)	Prec(%)	Rec(%)	F1-Sc(%)
2000	32x32	80,5	83	81	80
	128x128	80,5	84	81	80
6500	32x32	82,5	85	82	82
	128x128	82,5	85	82	82
10320	32x32	80,5	83	81	80
	128x128	81	84	81	81
12130	32x32	79,5	84	80	79
	128x128	79,5	84	80	79

Tablo 6.SVM algoritmalarının RGB renk formatında performans sonuçları

Veri Seti Boyutu	Görüntü Ölçeği	SVM			
		Acc(%)	Prec(%)	Rec(%)	F1-Sc(%)
2000	32x32	80,5	81	81	80
	128x128	89	89,5	88	87
6500	32x32	875	88	88	87
	128x128	84,5	85	84	84
10320	32x32	82,5	85	82	82
	128x128	79,5	82	80	79
12130	32x32	79,5	80	80	79
	128x128	76	80	76	75

Bu tablolara bakıldığında en iyi test sonucunun Tablo.6'daki 2000 veri seti, 128x128 görüntü ölçeği ve SVM

algoritması kullanılarak elde edildiği görülmektedir. Acurracy performans metriğine göre %89, Precision metriğine göre %89,5, Recall metriğine göre %88 ve F1-score metriğine göre %87 oranında başarı elde edilmiştir. ALL görüntülerinin sınıflandırılması sonucunda SVM algoritması benzer çalışmalarda da [5], [6] görüldüğü gibi daha başarılı olduğu görülmektedir.

IV. SONUÇ

Çalışmada veri seti boyutu, görüntü ölçeği ve renk formatı değiştirilerek ALL veri kümesi üzerindeki başarı değerlerinin karşılaştırılması yapılmıştır.

Elde edilen sonuçlara göre veri setinin büyüklüğünün başarıya katkıda bulunmadığı fakat görüntü ölçeğinin büyütülmesi ve RGB renk formatının kullanılmasının başarıya olumlu katkı yaptığı gözlenmiştir.

Aynı zamanda 2'li sınıflandırma yapabilmek için SVM algoritmasının çekirdek parametresinin '*linear'* olarak seçilmesinin daha iyi sonuç ürettiğini ve bu yüzden en iyi sonucu SVM algoritmasıyla elde ettiğimizi düşünmekteyiz.

Gelecekte sınıflandırma işlemlerini kovolüsyonel sinir ağları ile gerçekleştirip bu çalışmada kullanılan makine öğrenimi yöntemleri ile karşılaştırılması üzerine çalışmalar yapılacaktır.

TEŞEKKÜRLER

Cancer Imaging Archive tarafından erişim sağlanan ALL veri setinin kullanılmasına izin veren ve düzenlenmesine katkı sağlayan tüm yetkililere teşekkür ederiz [15], [16], [17], [18], [19], [20], [21].

KAYNAKLAR

[1] (2019) WHO website [Online]. Available: https://gco.iarc.fr/today/home

[2] T. Aksu Uzunhan, Z. Karalaş, "Çocukluk Çağı Akut Lenfoblastik Lösemisi," *Çocuk Dergisi* vol.12, no.1, pp. 6-15, 2012

[3] A. A. Atay, A. E. Kürekçi, V. Kesik, S. Kýlýç, M. Gülgün, O. Özcan ve E. Gökçay, "Akut lenfoblastik lösemili olgularımızın retrospektif analizi, "*Gülhane Tıp Dergisi, vol.* 47, pp. 183-186, 2005

[4] A. M. Abdeldaim, A. T. Sahlol, M. Elhoseny & A. E. Hassanien, *In Advances in Soft Computing and Machine Learning in Image Processing*, Computer-aided acute lymphoblastic leukemia diagnosis system based on image analysis. pp. 131-147, Springer, Cham, 2018

[5] M. N. Q. Bhuiyan, S. K. Rahut, R. A. Tanvir, & S.VRipon, "Automatic Acute Lymphoblastic Leukemia Detection and Comparative Analysis from Images," *In 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), IEEE,* 2019, pp. 1144-1149

[6] O. Selçuk & F. Özen, , "Acute lymphoblastic leukemia diagnosis using image processing techniques,". *In 2015 23nd Signal Processing and Communications Applications Conference (SIU),* IEEE, 2015 pp. 803-806

[7] (2019) NumPy website [Online]. Available: https://numpy.org/

[8] (2019) Matplotlib website [Online]. Available: https://matplotlib.org/

[9] (2019) Skit-Learn website [Online]. Available: https://scikit-learn.org/stable/

[10] R. Saravanan, & P. Sujatha, "A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification", *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, 2018, pp. 945-949

[11] (2017) M. N. Bernstein websites [Online]. Available: http://pages.cs.wisc.edu/~matthewb/pages/notes/pdf/ensembles/RandomForests.pdf

[12] R. Gandhi. (2018) Towards Data Science homepage, [Online]. Available: https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c

[13] U. Köse, "Zeki Optimizasyon Tabanlı Destek Vektör Makineleri ile Diyabet Teşhisi," *Politeknik Dergisi*, vol. 22, no.3, p.p. 557-566, 2019

[14] T. S Furey, N. Cristianini, N.Duffy, D. W. Bednarski, M. Schummer, D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data" , *Bioinformatics*, vol. 16, no. 10, pp. 906–914, Oct. 2000.

[15] A. Gupta, R. Gupta, "ALL Challenge dataset of ISBI 2019 [Data set]," *The Cancer Imaging Archive*, 2019

[16] A. Gupta, R. Duggal, R. Gupta, L. Kumar, N. Thakkar, and D. Satpathy, "GCTI-SN: Geometry-Inspired Chemical and Tissue Invariant Stain Normalization of Microscopic Medical Images,", under review.

[17] R. Gupta, P. Mallick, R. Duggal, A. Gupta, and O. Sharma, "Stain Color Normalization and Segmentation of Plasma Cells in Microscopic Images as a Prelude to Development of Computer Assisted Automated Disease Diagnostic Tool in Multiple Myeloma," *16th International Myeloma Workshop (IMW),* India, March 2017.

[18] R. Duggal, A. Gupta, R. Gupta, M. Wadhwa, and C. Ahuja, "Overlapping Cell Nuclei Segmentation in Microscopic Images UsingDeep Belief Networks," *Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, India, December 2016.

[19] R. Duggal, A. Gupta, and R. Gupta, "Segmentation of overlapping/touching white blood cell nuclei using artificial neural networks," *CME Series on Hemato-Oncopathology, All India Institute of Medical Sciences (AIIMS)*, New Delhi, India, July 2016.

[20] R. Duggal, A. Gupta, R. Gupta, and P. Mallick, "SD-Layer: Stain Deconvolutional Layer for CNNs in Medical Microscopic Imaging," In: Descoteaux M., Maier-Hein L., Franz A., Jannin P., Collins D., Duchesne S. (eds) Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017, MICCAI 2017. Lecture Notes in Computer Science, Part III, LNCS 10435, pp. 435–443. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-66179-7_50.

[21] TCIA Citation K. Clark, B. Vendt, K. Smith, J. Freymann , J. Kirby , P. Koppel , S. Moore , S. Phillips , D. Maffitt , M Pringle, L. Tarbox, F. Prior, " The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *Journal of Digital Imaging*, Volume 26, Number 6, December, 2013, pp 1045-1057.